

CMR Adapter Use Cases

THESE USE CASES HAVE BEEN MOVED TO JAMA FOR DISCUSSION AND REVIEW. PLEASE DON'T EDIT THIS PAGE. See the Pre-CMR Jama project for comments and discussion.

Use Case Review Link (please contact [Dan Pilone](#) if you don't have sufficient access).

<https://rms.earthdata.nasa.gov/review.req#/r:REV-17>

----- 8< -----

CMR Adapter Initial Requirements

Dan: Top level question: How will some of these Use Cases fit with Automated Provider Reconciliation? For example, ECS (SDPS/BMGT) providers regularly reconcile what's in ECHO (and presumably the CMR) against their internal holdings. Most of these uses cases will cause a metadata mismatch which will result in a re-export, which will be readapted, etc, etc....

===Use Case 1===

Title: Automatically Insert Specific Controlled Vocabulary into Metadata Records (DIF/ECHO/EMS).

Description: Automatically populate EOSDIS metadata collections with specific Controlled Vocabulary, such as a science keyword, project, platform, etc. For example, all NASA EOSDIS collections should be tagged with "Project=EOSDIS", and "Project=ESIP".

===Use Case 2===

Title: Check metadata for compliance with Required Fields (DIF, ECHO, EMS)

Description: There are cases where the required fields of DIF, ECHO, EMS, and the UMM are not met. Metadata records without the required fields should be flagged and populated with the required fields if possible.

===Use Case 3===

Title: Auto-insert DIF/Data Resolution Ranges

Description: There are cases in metadata records where Data Resolution units are populated, but not the range. Therefore the Data Resolution range would need to be auto-populated. This applies to Horizontal_Resolution_Range, Vertical_Resolution_Range, and Temporal_Resolution in the DIF format.

===Use Case 4===

Title: Verify and Map Controlled Vocabulary To Current Version

Description: Some data providers use older/previous controlled vocabulary versions. In this instance, the older keyword would need to be mapped and converted to the current keyword.

===Use Case 5===

Title: DIF/ECHO/EMS Fields Will Have to Map to Related Fields in UMM model

Description: If there is a field from ECHO 10 and DIF that needs to be made compliant with the UMM model, it will have to be adjusted to meet the requirements of the UMM.

===Use Case 6===

Title: Transfer Older Versions of DIF, ECHO to latest version

Description: A user may have a metadata record that uses an older version/schema of the DIF and/or ECHO, and needs to convert the metadata record to the most recent version of DIF or ECHO.

===Use Case 7===

Title: Metadata Translation from NASA ISO flavor to DIF (UMM), ECHO (UMM)

Description: A user may have a metadata record in NASA ISO flavor that needs to be converted to DIF (UMM) or ECHO (UMM).

Comments:

- Jason: How will we know if we need to translate a record from ISO to another format?
- Scott: I am not sure how we would automatically know that the record is a NASA flavored ISO. We have to check but there might be something in the XML header.
- Jason: Is this on a case by case basis?
- Scott: Not sure what you mean here, but if the incoming metadata is in ISO it will need to be translated to the requested UMM formatted form (DIF, ECHO).
- Jason: Why can't we leave the data in ISO format in the CMR? The CMR will have the ability to retrieve any record in the desired format of the user. (i.e. DIF, ECHO10, ISO)
- Scott: You cannot leave the document in ISO because there isn't a UMM compatible version of ISO. Also, it will be very unlikely that any two providers use ISO the same way. Therefore in order to maintain metadata consistency and quality in EOSDIS systems the adapters will need to translate the ISO metadata into a UMM compatible format DIF (UMM) or ECHO (UMM).
- Dan: I think there's some confusion here between what's happening where:
 - The adapters serve to validate and massage or inject data if necessary to make a record UMM compliant, not to transform formats during ingest.
 - Formats **cannot** be transformed during ingest as they are all lossy translations in some form or another. DIF -> ECHO, ECHO -> DIF, ECHO -> ISO, ISO -> DIF, etc. all have some amount of misalignment so *something* could get lost in those transformations.
 - As a consequence, the CMR **must** be able to persist the native format (possibly with additional information injected or keywords mapped, etc.) of a record when ingested.
 - Transformations **can** happen during retrieval if so requested by the end user. The end user must be able to determine the "native" format of a record, if they're so inclined, but can also say "I don't care what the native format was, I only speak DIF" and get all results back in DIF format. This request may result in an ISO->DIF or ECHO->DIF conversion and potential loss of information that would have been in the native format but not supported (and therefore obviously not a UMM required field) in the target format.
 - Since the CMR must be able to persist native formats for records, the UMM must be able to define "validity" for all supported formats. The UMM, since it's not a format, is effectively a definition of the required fields for validation, recommended fields, and searchable fields. There should then exist a mapping from the UMM list to each supported format.

===Use Case 8===

Title: Capture and Preserve metadata content Lost or Gained During Translations

Description: There may be cases where metadata content in DIF and ECHO 10 metadata formats may be lost or gained during the transition process and adapters will be need to be able to preserve that information.

Comments:

- Jason: Where does that lost information get preserved? Is it just moved to a different part of the target format? If that's the case I would suggest we make that preservation part of the actual normal translation process.
- Scott: I think we should talk about this in a meeting. One option (from my non-technical point of view) is to move the information that we know will be lost to an "Extended Metadata" field in the target format. Another, is to possible preserve the metadata in its native format and link it to the target metadata.
- Dan: I suggest that adapter's **can't** convert between formats during ingest for exactly this reason. (See my comments on UC7 for more details)

===Use Case 9===

Title: Remove Duplicate Keywords

Description: There may be cases where duplicate keywords have been added to the DIF and ECHO 10 records. These duplicate keywords need to be identified and removed before loading.

Comments:

- Dan: At what point do we push back on the provider to clean up the actual metadata being submitted? As in, at what point do we reject a record vs. patching bad metadata?

===Use Case 10===

Title: Reconcile and Merge Controlled Vocabulary Across DIF and ECHO 10

Description: Metadata records that are ingested in DIF and ECHO 10 may use a slightly different name scheme/keyword for the same information. These keywords will need to be merged into the authoritative keyword that will be used in the system (i.e. GCMD KMS).

===Use Case 11===

Title: Automatically remove controlled vocabulary or text from DIF/ECHO records

Description: In some instances, keywords or other text may need to be removed from incoming metadata records (data provider specific).

Comments:

- Dan: This is another one that really feels like it's crossing into the "Bad metadata" space...

===Use Case 12===

Title: Remove/replace illegal characters in DIF/ECHO fields

Description: Some data providers include illegal characters in fields, such as Entry ID. These illegal characters will either need to be removed or replaced before the metadata can be accepted into the system.

Comments:

- Jason: Is this done automatically for every provider? How would the providers be notified that they have illegal characters in that field so they can fix the issue?
- Scott: The metadata validator will automatically check each record for non UTF-8 characters and characters that are not permitted within a certain field. The validator will notify the provider that there is an issue and where it occurs so it can be fixed on the provider end. In the meantime the adapter should strip the character out. Ideally the provider will fix the characters before the metadata is next ingested. If the provider does not fix the characters, the adapter should continue to strip out the bad characters.
- Dan: Can this happen for unique identification type fields, like short name or dataset ID? It seems like you could run into issues with orphaned collections if the adapter makes a change to an identity type field.

===Use Case 13===

Title: Replace a URL with another URL (or sections of a URL)

Description: Data providers change the URL (or sections of a URL) for all of the links in their system. Therefore, the corresponding links in the metadata need to be changed.

Comments:

- Jason: This sounds like a temporary fix. How quickly does this change need to be put into place? What about the existing metadata already ingested that points to the wrong location?
- Scott: The adapter should replace the broken link with the good link at ingest. The metadata validator will notify the provider that there is a broken link. Ideally the provider will fix the link before the next metadata ingest. If the provider does not fix the link, the adapter should continue to make that URL replace at ingest. Broken links in existing metadata should be globally replaced with the good links.
- Dan: I suggest that link checking can't be done during ingest. Ingest performance requirements would likely preclude the ability to check all links in a record. Link checking can be done asynchronously, but then the adapters are out of the picture. This feels like it should be a metadata update (either by the provider or by Science Coordinators) to the metadata rather than something that's codified in an adapter.

===Use Case 14===

Title: Insert xml/text into a specific metadata field

Description: A simple example would be to insert a string of text into the Access Constraints field. A more complicated example would be: a data center offers an API that provides direct access to data in an ordering system through that API. Therefore, if that API loaded with the Entry does not exist in the Related_URL fields within a specific metadata set, retrieve the ID from DIF/Entry_ID field, append ID to the end of API URL, insert URL into field along with textual information.